第一章 条件期望、条件方差和线性投影

内容提要

条件期望在金融计量学中扮演着重要的角色，经典线性回归模型的核心估计方法最小二乘（Least squares）本质上是在估计以解释变量 $\mathbf{x}$ 为条件被解释变量 y 的条件期望。而且，根据 $\mathbf{x}$ 提供的信息，关于 y 的最好（最小化均方误差）的预测也正是条件期望 $\mathbb{E}(\mathrm{y}|\mathbf{x})$。因此，本章将对条件期望的定义及其性质做一个较为完整的梳理。同时，本章还将讨论条件方差以及条件期望的最优线性近似（也就是：线性投影）的定义和性质。想深入了解本章内容可以进一步参考 [1-3]。

1.1 条件期望

定义 1.1. 条件期望

令 $\mathbf{x}$ 为 K 维随机向量，y 为一元随机变量。以 $\mathbf{x}$ 为条件，y 的条件期望函数 (Conditional Expectation Function, CEF) 定义如下：

$$\mu(\mathbf{x}) = \mathbb{E}(\mathrm{y}|\mathbf{x}) = \begin{cases} \int_{\mathrm{y}} y \cdot f(y|\mathbf{x})dy, & \mathrm{y}\text{ 连续}, \\ \sum_{\mathrm{y}} y \cdot \mathbb{P}_{\mathrm{y}|\mathbf{x}}(y|\mathbf{x}), & \mathrm{y}\text{ 离散}, \end{cases}$$

其中，$f(y|\mathbf{x})$ 是以 $\mathbf{x}$ 为条件 $\mathrm{y}=y$ 的条件概率密度函数，$\mathbb{P}_{\mathrm{y}|\mathbf{x}}(y|\mathbf{x})$ 是以 $\mathbf{x}$ 为条件 $\mathrm{y}=y$ 的条件概率。♣

在定义1.1中，条件期望函数被记为 $\mu(\mathbf{x})$，这表明 CEF 是随机向量 $\mathbf{x}$ 的函数，与随机变量 y 的具体实现值无关。关于条件期望，我们有如下一些重要的结论：

定理 1.1. 简单迭代期望律 (Simple Law of Iterated Expectations, SLIE)

若 $\mathbb{E}|\mathrm{y}| < \infty$，则对任意的随机向量 $\mathbf{x}$，我们有：

$$\mathbb{E}_{\mathbf{x}}[\mathbb{E}[\mathrm{y}|\mathbf{x}]] = \mathbb{E}[\mathrm{y}],$$

其中，符号 $\mathbb{E}_{\mathbf{x}}[\cdot]$ 表示相对于随机向量 $\mathbf{x}$ 求期望。♡

证明 假设 y 和 $\mathbf{x}$ 的联合概率密度函数为 $f(y,\boldsymbol{x})$，因为 $\mathbb{E}[\mathrm{y}|\mathbf{x}]$ 是 $\mathbf{x}$ 的函数，我们有：

$$\mathbb{E}[\mathbb{E}[\mathrm{y}|\mathbf{x}]] = \int_{\mathbb{R}^k} \mathbb{E}[\mathrm{y}|\boldsymbol{x}] f_{\mathbf{x}}(\boldsymbol{x})d\boldsymbol{x}.$$

将 y 为连续随机变量情况下条件期望的定义1.1代入上式得到：

$$\int_{\mathbb{R}^k}\left(\int_{\mathbb{R}} yf(y|\boldsymbol{x})dy\right) f_{\mathbf{x}}(\boldsymbol{x})d\boldsymbol{x} = \int_{\mathbb{R}^k}\int_{\mathbb{R}} yf(y,\boldsymbol{x})dyd\boldsymbol{x} = \int_{\mathbb{R}} yf(y)dy = \mathbb{E}[\mathrm{y}],$$

其中，第一个等号成立是因为 $f(y|\boldsymbol{x})f_{\mathbf{x}}(\boldsymbol{x}) = f(y,\boldsymbol{x})$；第二个等号成立是因为联合分布和边缘分布之间存在如下关系：$\int_{\mathbb{R}^k} f(y,\boldsymbol{x})d\boldsymbol{x} = f(y)$。 □

为了更好地理解定理1.1，我们假设 $\mathbf{x}$ 是离散随机变量，则有：

$$\mathbb{E}[\mathbb{E}[\mathrm{y}|\mathbf{x}]] = \sum_{j=1}^{\infty} \mathbb{E}\left[\mathrm{y}|\mathbf{x}=\boldsymbol{x}_j\right]\mathbb{P}\left[\mathbf{x}=\boldsymbol{x}_j\right].$$

如果我们将求期望理解为求平均, 那么条件均值的加权（以 $\mathbf{x}$ 取不同值的概率作为权重）平均是无条件均值。例如，令 $\mathbf{x}$ 代表性别，y 代表身高，

$$\mathbb{E}[\text{身高} \mid \text{性别}=\text{女}]\mathbb{P}\left[\text{性别}=\text{女}\right] + \mathbb{E}[\text{身高} \mid \text{性别}=\text{男}]\mathbb{P}\left[\text{性别}=\text{男}\right] = \mathbb{E}[\text{身高}].$$

所以，如果我们想得到一个人口样本的平均身高，可以直接求平均身高，也可以分别先求得女性和男性的平均身高再以各个性别所占的比例作为权重取加权平均。

定理 1.2. 迭代期望律 (Law of Iterated Expectations, LIE)

若 $\mathbb{E}|\mathrm{y}| < \infty$，则对任意的随机向量 $\mathbf{x}_1$ 和 $\mathbf{x}_2$，我们有：

$$\mathbb{E}_{\mathbf{x}_1}[\mathbb{E}[\mathrm{y}|\mathbf{x}_1,\mathbf{x}_2]|\mathbf{x}_1] = \mathbb{E}[\mathrm{y}|\mathbf{x}_1].$$

♡

证明 为了简化证明，这里假设定理中所涉及的随机变量具有联合密度函数。注意到

$$f\left(y \mid \boldsymbol{x}_1,\boldsymbol{x}_2\right) f\left(\boldsymbol{x}_2 \mid \boldsymbol{x}_1\right) = \frac{f\left(y,\boldsymbol{x}_1,\boldsymbol{x}_2\right)}{f\left(\boldsymbol{x}_1,\boldsymbol{x}_2\right)}\frac{f\left(\boldsymbol{x}_1,\boldsymbol{x}_2\right)}{f\left(\boldsymbol{x}_1\right)} = f\left(y,\boldsymbol{x}_2 \mid \boldsymbol{x}_1\right), \tag{1.1}$$

以及，

$$\mathbb{E}[\mathrm{y}|\mathbf{x}_1=\boldsymbol{x}_1,\mathbf{x}_2=\boldsymbol{x}_2] = \int_{\mathbb{R}} y f(y \mid \boldsymbol{x}_1,\boldsymbol{x}_2)dy. \tag{1.2}$$

令随机向量 $\mathbf{x}_2$ 的维度为 k_2，于是，我们有：

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[\mathrm{y}|\mathbf{x}_1,\mathbf{x}_2]|\mathbf{x}_1=\boldsymbol{x}_1] &= \int_{\mathbb{R}^{k_2}} \mathbb{E}[\mathrm{y}|\mathbf{x}_1=\boldsymbol{x}_1,\mathbf{x}_2=\boldsymbol{x}_2]f(\boldsymbol{x}_2 \mid \boldsymbol{x}_1)d\boldsymbol{x}_2 \\
&= \int_{\mathbb{R}^{k_2}} \left(\int_{\mathbb{R}} y f(y \mid \boldsymbol{x}_1,\boldsymbol{x}_2)dy\right) f(\boldsymbol{x}_2 \mid \boldsymbol{x}_1)d\boldsymbol{x}_2 \\
&= \int_{\mathbb{R}^{k_2}}\int_{\mathbb{R}} y f(y \mid \boldsymbol{x}_1,\boldsymbol{x}_2)f(\boldsymbol{x}_2 \mid \boldsymbol{x}_1)dyd\boldsymbol{x}_2 \\
&= \int_{\mathbb{R}^{k_2}}\int_{\mathbb{R}} y f\left(y,\boldsymbol{x}_2 \mid \boldsymbol{x}_1\right)dyd\boldsymbol{x}_2 \\
&= \mathbb{E}[\mathrm{y}|\mathbf{x}_1=\boldsymbol{x}_1],
\end{aligned}$$

其中，第二个等式用到了式 (1.2)；第四个等式用到了式 (1.1)。

由上式易知，$\mathbb{E}[\mathbb{E}[\mathrm{y}|\mathbf{x}_1,\mathbf{x}_2]|\mathbf{x}_1] = \mathbb{E}[\mathrm{y}|\mathbf{x}_1]$。 □

为了便于记住定理1.2，我们可以把此性质总结为"小信息集占优"。此定理在时间序列分析中有广泛的应用，例如，令 $\mathscr{F}_t$ 表示 t 时刻所能获得的信息，我们有：若 $s<t$，$\mathbb{E}\{\mathbb{E}[\mathrm{y}_{t+h}|\mathscr{F}_t]|\mathscr{F}_s\} = \mathbb{E}[\mathrm{y}_{t+h}|\mathscr{F}_s]$。实际上，可以证明如下更一般形式的迭代期望律：令 $\mathbf{w}$ 为随机向量，$\mathbf{x}$ 是随机向量且满足 $\mathbf{x}=\mathbf{g}(\mathbf{w})$[1]。一般形式的 LIE 可表示为：

$$\mathbb{E}[\mathrm{y}|\mathbf{x}] = \mathbb{E}\left[\mathbb{E}(\mathrm{y}|\mathbf{w})|\mathbf{x}\right].$$

[1]如果已知 $\mathbf{w}$ 的实现值，$\mathbf{x}$ 的实现值便是已知的。

定理 1.3. 条件化定理 (Conditioning Theorem, CT)

若 $\mathbb{E}|\mathrm{y}| < \infty$，则，

$$\mathbb{E}[g(\mathbf{x}) \cdot \mathrm{y}|\mathbf{x}] = g(\mathbf{x}) \cdot \mathbb{E}[\mathrm{y}|\mathbf{x}].$$

若再假设 $\mathbb{E}|g(\mathbf{x})| < \infty$，则，

$$\mathbb{E}[g(\mathbf{x}) \cdot \mathrm{y}] = \mathbb{E}[g(\mathbf{x}) \cdot \mathbb{E}[\mathrm{y}|\mathbf{x}]].$$

♡

证明 给定 $\mathbf{x} = \boldsymbol{x}$，我们有：

$$\mathbb{E}[g(\mathbf{x}) \cdot \mathrm{y}|\mathbf{x} = \boldsymbol{x}] = \int_{\mathbb{R}} g(\boldsymbol{x}) y f(y|\boldsymbol{x}) dy = g(\boldsymbol{x}) \int_{\mathbb{R}} y f(y|\boldsymbol{x}) dy = g(\boldsymbol{x}) \mathbb{E}(\mathrm{y}|\mathbf{x} = \boldsymbol{x}).$$

因此，$\mathbb{E}[g(\mathbf{x}) \cdot \mathrm{y}|\mathbf{x}] = g(\mathbf{x}) \cdot \mathbb{E}[\mathrm{y}|\mathbf{x}]$。由定理1.1，本定理中的第二个式子容易验证。 □

下面我们给出关于条件期望函数的两个重要性质：

性质 （CEF 分解性质）对任意的随机变量 y，有如下性质成立：

$$\mathrm{y} = \mathbb{E}[\mathrm{y}|\mathbf{x}] + \varepsilon,$$

其中，

(a) ε 均值独立于 $\mathbf{x}$，即，$\mathbb{E}[\varepsilon|\mathbf{x}] = 0$;

(b) ε 与 $\mathbf{x}$ 的任意函数都不相关。

证明 首先证明 (a),

$$\mathbb{E}[\varepsilon|\mathbf{x}] = \mathbb{E}\{\mathrm{y} - \mathbb{E}[\mathrm{y}|\mathbf{x}]|\mathbf{x}\} = \mathbb{E}[\mathrm{y}|\mathbf{x}] - \mathbb{E}[\mathrm{y}|\mathbf{x}] = 0.$$

其次证明 (b)，假设 $h(\mathbf{x})$ 是 $\mathbf{x}$ 的任意函数，我们有：

$$\begin{aligned}
\mathrm{Cov}[\varepsilon, h(\mathbf{x})] &= \mathbb{E}[\varepsilon \cdot h(\mathbf{x})] - \mathbb{E}(\varepsilon)\mathbb{E}[h(\mathbf{x})] \\
&= \mathbb{E}\{\mathbb{E}[\varepsilon \cdot h(\mathbf{x})|\mathbf{x}]\} - 0 \cdot \mathbb{E}[h(\mathbf{x})] && \text{(这里使用了两次定理1.1)} \\
&= \mathbb{E}\{h(\mathbf{x})\mathbb{E}[\varepsilon|\mathbf{x}]\} && \text{(这里使用了定理1.3)} \\
&= 0. && \text{(这里使用了 (a) 的结论)}
\end{aligned}$$

□

性质 （CEF 预测性质）假设 $\mathbb{E}(\mathrm{y}^2) < \infty$，记 $\mu(\mathbf{x}) \equiv \mathbb{E}(\mathrm{y}|\mathbf{x})$，则 μ 是如下优化问题的解:

$$\min_{m \in \mathscr{M}} \mathbb{E}[(\mathrm{y} - m(\mathbf{x}))^2],$$

其中，$\mathscr{M}$ 是函数类，$m : \mathbb{R}^K \to \mathbb{R}$ 满足 $\mathbb{E}[m(\mathbf{x})^2] < \infty$。

证明 首先，我们对优化的目标函数进行分解：

$$\begin{aligned}
& \mathbb{E}[(\mathrm{y} - m(\mathbf{x}))^2] \\
= \ & \mathbb{E}[(\mathrm{y} - \mu(\mathbf{x}) + \mu(\mathbf{x}) - m(\mathbf{x}))^2] \\
= \ & \mathbb{E}[(\mathrm{y} - \mu(\mathbf{x}))^2] + \mathbb{E}[(\mu(\mathbf{x}) - m(\mathbf{x}))^2] + 2 \cdot \mathbb{E}\{[\mathrm{y} - \mu(\mathbf{x})] \cdot [\mu(\mathbf{x}) - m(\mathbf{x})]\}.
\end{aligned}$$

注意到，我们需要从函数类 $\mathscr{M}$ 中选择函数 m 使目标函数最小化，上述分解中的第一项与 m 的选择无关，第二项在 $m(\mathbf{x}) = \mu(\mathbf{x})$ 时取到最小值 0，我们只需证明第三项的取值与 m 的选择无关，实际上，

$$\mathbb{E}\{[\mathrm{y} - \mu(\mathbf{x})] \cdot [\mu(\mathbf{x}) - m(\mathbf{x})]\} = \mathbb{E}\{\varepsilon \cdot [\mu(\mathbf{x}) - m(\mathbf{x})]\} = 0,$$

其中，第二个等号成立是因为 $\mu(\mathbf{x})-m(\mathbf{x})$ 是 $\mathbf{x}$ 的函数，利用 CEF 分解性质 (b) 即可。

于是，目标函数的最小值在 $m(\mathbf{x})=\mu(\mathbf{x})$ 时取到，此性质得证。 □

对 $\mathbf{x}$ 任意的函数 $m(\mathbf{x})$，定义其预测 y 得到的均方误差 (mean squared error, MSE)为

$$\mathrm{MSE}(\mathrm{y};m)\equiv\mathbb{E}[(\mathrm{y}-m(\mathbf{x}))^2],$$

CEF 的分解性质表明：对任意的随机变量 y 和随机向量 $\mathbf{x}$，y 总是可以分解为 $\mathbf{x}$ 可以解释的部分 $\mathbb{E}[\mathrm{y}|\mathbf{x}]$ 和与 $\mathbf{x}$ 不相关的部分 ε。对于 $\mathbf{x}$ 可以解释的部分 $\mathbb{E}[\mathrm{y}|\mathbf{x}]$，CEF 的预测性质指出它是关于 y 的最小化均方误差的预测。金融计量学的重要任务之一就是研究如何基于已有的信息对金融时间序列进行预测，因此，条件期望是金融计量学重要的研究对象。值得一提的是，传统的金融计量方法通常假设条件期望函数具有线性函数形式，但实际上 $\mathbb{E}[\mathrm{y}|\mathbf{x}]$ 可以是 $\mathbf{x}$ 的非线性函数。为此，金融计量学也发展出诸如参数非线性模型、非参数或半参数模型、神经网络模型等一系列非线性模型。另外，金融计量学还包括公司金融实证研究使用的计量方法，特别是关于因果推断或政策评估的计量方法，这些方法关注的重点不是条件期望，而是一些因果效应参数的估计。

1.2 条件方差

定义 1.2. 条件方差

令 $\mathbf{x}$ 为随机向量，y 为一元随机变量。如果 $\mathbb{E}[\mathrm{y}^2]<\infty$，以 $\mathbf{x}=\boldsymbol{x}$ 为条件，y 的条件方差 (Conditional Variance) 定义如下：

$$\sigma^2(\boldsymbol{x})=\mathrm{Var}[\mathrm{y}|\mathbf{x}=\boldsymbol{x}]=\mathbb{E}\left[(\mathrm{y}-\mathbb{E}[\mathrm{y}|\mathbf{x}=\boldsymbol{x}])^2\,|\mathbf{x}=\boldsymbol{x}\right].$$

条件方差作为随机变量 $\mathbf{x}$ 的函数可记为 $\sigma^2(\mathbf{x})$。♣

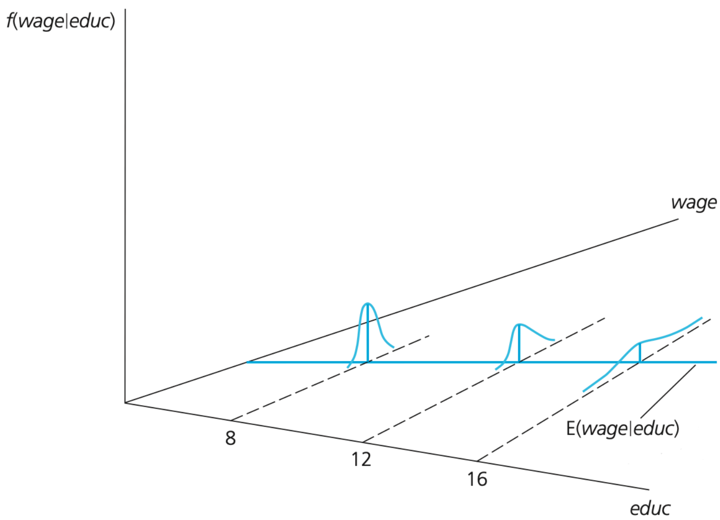

图 1.1: 条件期望与条件方差

由条件方差的定义可知条件方差也是随机变量 $\mathbf{x}$ 的函数。给定 $\mathbf{x}$ 的具体实现值 $\boldsymbol{x}$，条件期望 $\mathbb{E}[\mathrm{y}|\mathbf{x}=\boldsymbol{x}]$ 计算的是 y 的条件均值，而条件方差 $\mathrm{Var}[\mathrm{y}|\mathbf{x}=\boldsymbol{x}]$ 度量的是 y 的条件分布与其条件均值之间的偏离程度。对于对称的条件分布，条件期望给出了条件分布

的中心位置，条件方差刻画的是条件分布偏离中心位置的程度。在图1.1中，$\mathbf{x}$ 表示受教育年限 $educ$，y 表示工资收入 $wage$，图中分别给出了受教育年限等于 8、12 和 16 年的条件下，工资收入条件分布的均值位置（三条虚线和条件均值曲线 $\mathbb{E}[wage|educ]$ 的交点）以及条件分布的密度函数。我们发现，受教育年限越高，工资收入偏离平均工资收入的程度越大，这种现象在金融计量学中被称为条件异方差 (Conditional Heteroskedasticity)。

下面我们给出一些关于条件方差的重要结论：

定理 1.4. 方差分解 (Decomposition of Variance, DV)

若 $\mathbb{E}[\mathrm{y}^2] < \infty$，则有:

$$\mathrm{Var}[\mathrm{y}] = \mathbb{E}\left[\mathrm{Var}\left[\mathrm{y}|\mathbf{x}\right]\right] + \mathrm{Var}\left[\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]\right].$$

证明 根据 CEF 分解性质，我们有：

$$\begin{aligned}
\mathrm{Var}[\mathrm{y}] &= \mathrm{Var}\left\{\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right] + \varepsilon\right\} \\
&= \mathrm{Var}[\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]] + \mathrm{Var}(\varepsilon) && (\varepsilon \text{ 和 } \mathbb{E}[\mathrm{y}|\mathbf{x}] \text{ 不相关}) \\
&= \mathrm{Var}[\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]] + \mathbb{E}(\varepsilon^2) && (\mathbb{E}(\varepsilon) = 0) \\
&= \mathrm{Var}[\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]] + \mathbb{E}[\mathbb{E}[\varepsilon^2|\mathbf{x}]] && (\text{由 SLIE}) \\
&= \mathrm{Var}[\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]] + \mathbb{E}\left[\mathbb{E}\left[\left[\mathrm{y} - \mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]\right]^2|\mathbf{x}\right]\right] && (\varepsilon = \mathrm{y} - \mathbb{E}[\mathrm{y}|\mathbf{x}]) \\
&= \mathrm{Var}[\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right]] + \mathbb{E}\left[\mathrm{Var}\left[\mathrm{y}|\mathbf{x}\right]\right]. && (\text{根据条件方差的定义})
\end{aligned}$$

□

方差分解定理表明 y 的方差可以分解成 $\mathbb{E}[\mathrm{y}|\mathbf{x}]$ 的方差与 ε 的方差之和，前者是可由 $\mathbf{x}$ 解释的部分，后者是与 $\mathbf{x}$ 无关的部分。方差分解定理可以推广到 y 的条件方差，我们有如下结论：

定理 1.5. 条件方差分解 (Decomposition of Conditional Variance, DCV)

若 $\mathbb{E}[\mathrm{y}^2] < \infty$，则有:

$$\mathrm{Var}(\mathrm{y}|\mathbf{x}) = \mathbb{E}[\mathrm{Var}(\mathrm{y}|\mathbf{x},\mathbf{z})|\mathbf{x}] + \mathrm{Var}[\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})|\mathbf{x}].$$

证明 首先，我们对 $\mathrm{Var}(\mathrm{y}|\mathbf{x})$ 进行分解：

$$\begin{aligned}
\mathrm{Var}(\mathrm{y}|\mathbf{x}) &\equiv \mathbb{E}\left\{[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x})]^2|\mathbf{x}\right\} \\
&= \mathbb{E}\left\{[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z}) + \mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z}) - \mathbb{E}(\mathrm{y}|\mathbf{x})]^2|\mathbf{x}\right\} \\
&= \mathbb{E}\left\{[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})]^2|\mathbf{x}\right\} + \mathbb{E}\left\{[\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z}) - \mathbb{E}(\mathrm{y}|\mathbf{x})]^2|\mathbf{x}\right\} \\
&\quad + 2\underbrace{\mathbb{E}\left\{[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})] \cdot [\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z}) - \mathbb{E}(\mathrm{y}|\mathbf{x})]|\mathbf{x}\right\}}_{=0}.
\end{aligned}$$

由 LIE 和条件方差的定义，该分解中的第一项可表示为：

$$\mathbb{E}\left\{[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})]^2|\mathbf{x}\right\} = \mathbb{E}\left\{\mathbb{E}\left\{[\mathrm{y} - \mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})]^2|\mathbf{x},\mathbf{z}\right\}|\mathbf{x}\right\} = \mathbb{E}[\mathrm{Var}(\mathrm{y}|\mathbf{x},\mathbf{z})|\mathbf{x}].$$

同理，第二项可表示为：

$$\mathbb{E}\left\{[\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z}) - \mathbb{E}(\mathrm{y}|\mathbf{x})]^2|\mathbf{x}\right\} = \mathbb{E}\left\{[\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z}) - \mathbb{E}(\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})|\mathbf{x})]^2|\mathbf{x}\right\} = \mathrm{Var}[\mathbb{E}(\mathrm{y}|\mathbf{x},\mathbf{z})|\mathbf{x}].$$

第三项为 0 的推导留作习题，见习题1.。 □

我们已经知道 CEF 是关于 y 的最小化均方误差的预测，若预测时我们使用更多的信息 $(\mathbf{x}, \mathbf{z})$，相比于 $\mathbb{E}[y|\mathbf{x}]$，$\mathbb{E}[y|\mathbf{x}, \mathbf{z}]$ 是否是关于 y 的更好的预测呢？由定理1.5，我们容易得到：

$$\mathbb{E}[\mathrm{Var}(y|\mathbf{x})] \geq \mathbb{E}[\mathrm{Var}(y|\mathbf{x}, \mathbf{z})],$$

其中，

$$\mathbb{E}[\mathrm{Var}(y|\mathbf{x})] = \mathbb{E}\left[\mathbb{E}\left[(y - \mathbb{E}[y|\mathbf{x}])^2 |\mathbf{x}\right]\right] = \mathbb{E}\left[(y - \mathbb{E}[y|\mathbf{x}])^2\right] = \mathrm{MSE}(y; \mathbb{E}[y|\mathbf{x}]),$$

同理，

$$\mathbb{E}[\mathrm{Var}(y|\mathbf{x}, \mathbf{z})] = \mathrm{MSE}(y; \mathbb{E}[y|\mathbf{x}, \mathbf{z}]).$$

于是，我们有：

$$\mathrm{MSE}(y; \mathbb{E}[y|\mathbf{x}]) \geq \mathrm{MSE}(y; \mathbb{E}[y|\mathbf{x}, \mathbf{z}]).$$

从上述分析我们得到 $\mathbb{E}[y|\mathbf{x}, \mathbf{z}]$ 的 MSE 不大于 $\mathbb{E}[y|\mathbf{x}]$ 的 MSE，也就是说，从总体 (population) 的角度讲，基于更多的信息对 y 进行预测有助于减少预测的均方误差。除此之外，上述分析与如下定理的结论等价：

定理 1.6. 方差缩减定理 (Variance Reduction Theorem, VRT)

若 $\mathbb{E}[y^2] < \infty$，则有：

$$\mathrm{Var}[y] \geq \mathrm{Var}[y - \mathbb{E}[y|\mathbf{x}]] \geq \mathrm{Var}[y - \mathbb{E}[y|\mathbf{x}, \mathbf{z}]].$$

♡

证明 记 $\varepsilon = y - \mathbb{E}[y|\mathbf{x}]$，则有：

$$\mathrm{Var}[y - \mathbb{E}[y|\mathbf{x}]] = \mathrm{Var}(\varepsilon) = \mathbb{E}(\varepsilon^2) = \mathrm{MSE}(y; \mathbb{E}[y|\mathbf{x}]).$$

同理，

$$\mathrm{Var}[y - \mathbb{E}[y|\mathbf{x}, \mathbf{z}]] = \mathrm{MSE}(y; \mathbb{E}[y|\mathbf{x}, \mathbf{z}]).$$

因此，第二个“$\geq$”得证。由定理1.4的证明易知 $\mathrm{Var}(y) \geq \mathrm{Var}(\varepsilon)$，第一个“$\geq$”得证。 □

此定理的结论实际上与我们在下一章定义的拟合优度和 R^2 的概念相对应，不同的是我们这里讨论的是总体的性质，下一章讨论的是样本的性质。

1.3 线性投影

虽然条件期望 $\mu(\mathbf{x})$ 提供了 y 的最优预测，但是 $\mu(\mathbf{x})$ 的函数形式通常是未知的。在实证分析时，一般假设 CEF 具有线性形式，但是这种假设往往不切实际。更准确的做法是将线性形式的函数设定看作是关于 CEF 的近似。为此，我们研究 CEF 的最优线性近似，即，具有最小均方误差的线性近似。

定义 1.3. 线性投影

给定 $\mathbf{x}$，y 的最优线性预测，也称为线性投影，被定义为：

$$\mathscr{P}(\mathrm{y}|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta},$$

其中，$\boldsymbol{\beta}$ 被称为线性投影系数，它最小化如下均方预测误差：

$$S(\boldsymbol{b}) = \mathbb{E}\left((\mathrm{y} - \mathbf{x}'\boldsymbol{b})^2\right).$$

即，

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{b}\in\mathbb{R}^K} S(\boldsymbol{b}).$$

♣

为了推导 $\boldsymbol{\beta}$ 的解析表达式，需要施加如下假设：

假设 1.1.

(a) $\mathbb{E}[\mathrm{y}^2] < \infty$;

(b) $\mathbb{E}\|\mathbf{x}\|^2 < \infty$;

(c) $\boldsymbol{Q}_{\mathbf{xx}} = \mathbb{E}[\mathbf{xx}']$ 是正定矩阵。

定理 1.7. 线性投影的性质

(1) 线性投影系数 $\boldsymbol{\beta}$ 存在且唯一，其表达式为：

$$\boldsymbol{\beta} = (\mathbb{E}[\mathbf{xx}'])^{-1}\mathbb{E}[\mathbf{x}\mathrm{y}].$$

(2) $\mathscr{P}(\mathrm{y}|\mathbf{x}) = \mathbf{x}'(\mathbb{E}[\mathbf{xx}'])^{-1}\mathbb{E}[\mathbf{x}\mathrm{y}]$.

(3) 投影误差 $e = \mathrm{y} - \mathbf{x}'\boldsymbol{\beta}$ 存在，并且满足 $\mathbb{E}[e^2] < \infty$ 和 $\mathbb{E}[\mathbf{x}e] = \mathbf{0}$。特别地，若 $\mathbf{x}$ 中含有常数项，则 $\mathbb{E}(e) = 0$。

♡

证明 首先，将 $S(\boldsymbol{b})$ 进行分解：

$$S(\boldsymbol{b}) = \mathbb{E}[\mathrm{y}^2] + \boldsymbol{b}'\mathbb{E}[\mathbf{xx}']\boldsymbol{b} - 2\boldsymbol{b}'\mathbb{E}[\mathbf{x}\mathrm{y}].$$

对上式两端取偏导，由最优化的一阶条件，我们有：

$$\mathbf{0} = \frac{\partial S(\boldsymbol{b})}{\partial \boldsymbol{b}}\Big|_{\boldsymbol{b}=\boldsymbol{\beta}} = -2\mathbb{E}[\mathbf{x}\mathrm{y}] + 2\mathbb{E}[\mathbf{xx}']\boldsymbol{\beta}.$$

移项整理得：

$$\mathbb{E}[\mathbf{xx}']\boldsymbol{\beta} = \mathbb{E}[\mathbf{x}\mathrm{y}]. \tag{1.3}$$

由 ?? 和假设 1.1 (b)，我们有：

$$\|\mathbb{E}[\mathbf{xx}']\| \leq \mathbb{E}\|\mathbf{xx}'\| = \mathbb{E}\|\mathbf{x}\|^2 < \infty.$$

同理，由 ??、柯西-施瓦茨不等式和假设 1.1，我们有：

$$\|\mathbb{E}[\mathbf{x}\mathrm{y}]\| \leq \mathbb{E}\|\mathbf{x}\mathrm{y}\| \leq \left(\mathbb{E}\|\mathbf{x}\|^2\right)^{\frac{1}{2}}\left(\mathbb{E}[\mathrm{y}^2]\right)^{\frac{1}{2}} < \infty.$$

因此，$\mathbb{E}[\mathbf{xx}']$ 和 $\mathbb{E}[\mathbf{x}\mathrm{y}]$ 是良定义的 (well defined)。再由假设 1.1 (c) 知，$\mathbb{E}[\mathbf{xx}']$ 可逆，则有：

$$\boldsymbol{\beta} = (\mathbb{E}[\mathbf{xx}'])^{-1}\mathbb{E}[\mathbf{x}\mathrm{y}]. \tag{1.4}$$

由定义1.3，(2) 显然成立。

最后，我们证明 (3)：

$$
\begin{aligned}
\mathbb{E}(e^2) &= \mathbb{E}\left[(\mathrm{y}-\mathbf{x}'\boldsymbol{\beta})^2\right] \\
&= \mathbb{E}[\mathrm{y}^2] - 2\mathbb{E}[\mathrm{y}\mathbf{x}']\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbb{E}[\mathbf{x}\mathbf{x}']\boldsymbol{\beta} \\
&= \mathbb{E}[\mathrm{y}^2] - \mathbb{E}[\mathrm{y}\mathbf{x}'](\mathbb{E}[\mathbf{x}\mathbf{x}'])^{-1}\mathbb{E}[\mathbf{x}\mathrm{y}] && \text{(将式 (1.4) 的结论代入)} \\
&\le \mathbb{E}[\mathrm{y}^2] < \infty. && \text{((}\mathbb{E}[\mathbf{x}\mathbf{x}'])^{-1}\text{ 半正定)}
\end{aligned}
$$

由式 (1.3)，我们有：

$$
\mathbf{0} = \mathbb{E}[\mathbf{x}\mathrm{y}] - \mathbb{E}[\mathbf{x}\mathbf{x}']\boldsymbol{\beta} = \mathbb{E}\left[\mathbf{x}\left(\mathrm{y}-\mathbf{x}'\boldsymbol{\beta}\right)\right] = \mathbb{E}\left[\mathbf{x}e\right].
$$

□

定理1.7给出了线性投影系数 $\boldsymbol{\beta}$ 的一般表达式，接下来我们考虑两种特殊情况：

- $\mathbf{x} = (1, \mathrm{x}_1)'$，在此种情况下，我们有：

$$
\mathrm{y} = \alpha + \beta_1 \mathrm{x}_1 + e,
$$

其中，α 和 β_1 分别是常数项 1 和 x_1 对应的线性投影系数。容易证明：

$$
\beta_1 = \frac{\mathrm{Cov}(\mathrm{y}, \mathrm{x}_1)}{\mathrm{Var}(\mathrm{x}_1)}, \quad \alpha = \mathbb{E}(\mathrm{y}) - \beta_1 \mathbb{E}(\mathrm{x}_1). \tag{1.5}
$$

- $\mathbf{x} = (1, \mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_K)'$，$K > 1$，在此种情况下，我们有：

$$
\beta_k = \frac{\mathrm{Cov}(\mathrm{y}, \tilde{\mathrm{x}}_k)}{\mathrm{Var}(\tilde{\mathrm{x}}_k)}, \tag{1.6}
$$

其中，β_k 是 x_k 对应的线性投影系数，$\tilde{\mathrm{x}}_k$ 是 x_k 对所有其他协变量投影得到的投影误差。下面我们对式 (1.6) 进行证明。

证明 首先，将 y 写成线性投影加投影误差的形式：

$$
\mathrm{y} = \alpha + \beta_1 \mathrm{x}_1 + \cdots + \beta_k \mathrm{x}_k + \cdots + \beta_K \mathrm{x}_K + e.
$$

将上式代入 $\mathrm{Cov}(\mathrm{y}, \tilde{\mathrm{x}}_k)$，注意到，由习题4.，我们有：

- $\mathrm{Cov}(\tilde{\mathrm{x}}_k, e) = 0$，因为 $\tilde{\mathrm{x}}_k$ 是 $\mathbf{x}$ 各个分量的线性组合，e 与 $\mathbf{x}$ 不相关；
- $\mathrm{Cov}(\tilde{\mathrm{x}}_k, \mathrm{x}_j) = 0$，$j \neq k$，因为 $\tilde{\mathrm{x}}_k$ 是线性投影误差，与 x_j 不相关；
- $\mathrm{Cov}(\tilde{\mathrm{x}}_k, \mathrm{x}_k) = \mathrm{Cov}(\tilde{\mathrm{x}}_k, \tilde{\mathrm{x}}_k + \mathbf{x}'_{-k}\boldsymbol{\beta}_{-k}) = \mathrm{Var}(\tilde{\mathrm{x}}_k)$，其中，$\mathbf{x}_{-k}$ 表示除 x_k 之外的所有协变量构成的向量，即，$\mathbf{x}_{-k} = (1, \mathrm{x}_1, \cdots, \mathrm{x}_{k-1}, \mathrm{x}_{k+1}, \cdots, \mathrm{x}_K)'$，$\boldsymbol{\beta}_{-k}$ 是 x_k 对 $\mathbf{x}_{-k}$ 进行投影得到的线性投影系数。

于是，

$$
\frac{\mathrm{Cov}(\mathrm{y}, \tilde{\mathrm{x}}_k)}{\mathrm{Var}(\tilde{\mathrm{x}}_k)} = \frac{\beta_k \mathrm{Var}(\tilde{\mathrm{x}}_k)}{\mathrm{Var}(\tilde{\mathrm{x}}_k)} = \beta_k.
$$

□

我们把第一种情况称为一元线性投影或简单线性投影，把第二种情况称为多元线性投影。由定理1.7中 (3) 的结论，线性投影误差 $\tilde{\mathrm{x}}_k$ 与 $\mathbf{x}_{-k}$ 正交，即，$\mathbb{E}[\mathbf{x}_{-k}\tilde{\mathrm{x}}_k] = \mathbf{0}$。式 (1.6) 说明多元线性投影中 x_k 的线性投影系数可以通过如下方式得到：首先，将 x_k 对 $\mathbf{x}_{-k}$ 做投影并保留误差 $\tilde{\mathrm{x}}_k$，目的是从 x_k 中剔除所有其他协变量的影响；然后，将 y 对误差 $\tilde{\mathrm{x}}_k$ 和常数项 1 做简单线性投影得到系数 β_k。

金融计量学（特别是应用于公司金融等实证方向的微观金融计量方法）的主要目的

之一是推断变量之间的因果关系。建立因果关系的核心是"控制其他所有相关因素不变"，拉丁语为 ceteris paribus。我们将在下一章多元线性回归中看到"控制其他所有相关因素不变"和"剔除所有其他因素的影响"是一致的。式 (1.6) 是从总体的角度讨论线性投影系数，下一章将从样本的角度讨论线性投影系数的估计。式 (1.6) 对应着多元线性回归的 Frisch-Waugh 定理。

关于线性投影，我们还有如下重要的结论：

定理 1.8. 线性投影与 CEF

(1) 假设 CEF 是 $\mathbf{x}$ 的线性函数，则 $\mu(\mathbf{x}) = \mathscr{P}(\mathrm{y}|\mathbf{x})$.

(2) 给定 $\mathbf{x}$，$\mathscr{P}(\mathrm{y}|\mathbf{x})$ 是关于 y 的最小化 MSE 的最优线性预测。

(3) $\mathscr{P}(\mathrm{y}|\mathbf{x})$ 是 $\mathbb{E}[\mathrm{y}|\mathbf{x}]$ 的最优（最小化 MSE）线性近似，即，

$$\boldsymbol{\beta} = \arg\min_{\boldsymbol{b}} \mathbb{E}\left\{\left(\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right] - \mathbf{x}'\boldsymbol{b}\right)^2\right\}.$$

♡

证明 首先，证明 (1)，假设 $\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}^*$，由 CEF 分解性质，$\mathbb{E}[\mathbf{x}(\mathrm{y} - \mathbb{E}[\mathrm{y}|\mathbf{x}])] = \mathbf{0}$，将 $\mu(\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}^*$ 代入并将 $\boldsymbol{\beta}^*$ 解出，得到 $\boldsymbol{\beta}^* = \boldsymbol{\beta}$。

由定义1.3知 (2) 显然成立。最后，我们证明 (3)，注意到：

$$\begin{aligned}(\mathrm{y} - \mathbf{x}'\boldsymbol{b})^2 &= \left\{(\mathrm{y} - \mathbb{E}[\mathrm{y}|\mathbf{x}]) + (\mathbb{E}[\mathrm{y}|\mathbf{x}] - \mathbf{x}'\boldsymbol{b})\right\}^2 \\ &= (\mathrm{y} - \mathbb{E}\left[\mathrm{y}|\mathbf{x}\right])^2 + \left(\mathbb{E}[\mathrm{y}|\mathbf{x}] - \mathbf{x}'\boldsymbol{b}\right)^2 + 2\left(\mathrm{y} - \mathbb{E}[\mathrm{y}|\mathbf{x}]\right)\cdot\left(\mathbb{E}[\mathrm{y}|\mathbf{x}] - \mathbf{x}'\boldsymbol{b}\right).\end{aligned}$$

所以，最小化 $\mathbb{E}\left\{(\mathbb{E}\left[\mathrm{y}|\mathbf{x}\right] - \mathbf{x}'\boldsymbol{b})^2\right\}$ 等价于最小化 $\mathbb{E}\left\{(\mathrm{y} - \mathbf{x}'\boldsymbol{b})^2\right\}$，由 (2) 知，(3) 成立。□

我们在前文中多次提到 CEF 提供了关于 y 的最优预测，而定理1.8的结论 (2) 说明线性投影提供了关于 y 的最优线性预测。除此之外，由于 CEF 的形式是未知的，我们一般假设其具有线性的形式，从定理1.8的结论 (3) 中可以看出这种假设具有一定的合理性。线性投影系数 $\boldsymbol{\beta}$ 也被称为总体回归系数 (population regression coefficient)，这是因为线性回归和线性投影具有密不可分的关系。实际上，线性回归是利用样本估计线性投影系数。

1.4 饱和回归模型

饱和回归模型是具有离散解释变量的回归模型，对解释变量取值的每一个可能组合，饱和回归模型都为该组合对应的条件期望引入一个额外的参数。我们将要看到饱和回归模型具有线性的 CEF。定理1.8的结论 (1) 表明：如果 CEF 具有线性形式，则 CEF 等价于线性投影。因此，对于饱和回归模型，我们有：$\mu(\mathbf{x}) = \mathscr{P}(\mathrm{y}|\mathbf{x})$。

首先，我们考虑只含有一个二值 (binary) 解释变量的情况，二值变量又被称为虚拟 (dummy) 变量。假设二值变量 x 代表性别，CEF 只能有两个不同的取值：

$$\mathbb{E}[\,\mathrm{y} \mid 性别\,] = \begin{cases} \mu_0 & 如果 \quad 性别 = 男性, \\ \mu_1 & 如果 \quad 性别 = 女性. \end{cases}$$

为了便于数学上的处理，我们引入虚拟变量 x_1，满足：

$$\mathrm{x}_1 = \begin{cases} 0 & 如果 \quad 性别 = 男性, \\ 1 & 如果 \quad 性别 = 女性. \end{cases}$$

于是，$\mathbb{E}[y|x_1] = \beta_0 + \beta_1 x_1$，其中，$\beta_0 = \mu_0$，$\beta_1 = \mu_1 - \mu_0$。此时，饱和回归模型可以表示为：

$$y = \beta_0 + \beta_1 x_1 + \varepsilon.$$

我们通常称 β_0 为截距，它表示男性总体 y 的期望；称 β_1 为斜率，它表示女性总体和男性总体 y 的期望之差。

类似地，若解释变量是多值离散变量 $x_1 \in \{0, 1, 2, \cdots, \tau\}$，关于 x_1 的饱和回归模型可以表示为：

$$y = \beta_0 + \beta_1 d_{11} + \beta_2 d_{12} + \cdots + \beta_\tau d_{1\tau} + \varepsilon,$$

其中，$d_{1j} = \mathbf{1}[x_1 = j]$，$j = 1, 2, \cdots, \tau$，$\mathbf{1}[A]$ 是示性函数，条件 A 成立时取值为 1，否则取值为 0。由上式容易验证：

$$\beta_0 = \mathbb{E}[y|x_1 = 0], \quad \beta_j = \mathbb{E}[y|x_1 = j] - \mathbb{E}[y|x_1 = 0].$$

接下来，我们考虑含有两个虚拟解释变量的情况。假设虚拟变量 x_2 满足：

$$x_2 = \begin{cases} 1 & \text{如果} \quad \text{具有大学学历,} \\ 0 & \text{如果} \quad \text{没有大学学历.} \end{cases}$$

以 x_1 和 x_2 为条件的条件期望函数有四种可能的取值：

$$\mathbb{E}[y|x_1, x_2] = \begin{cases} \mu_{00} & \text{如果} \quad x_1 = 0, x_2 = 0 \\ \mu_{01} & \text{如果} \quad x_1 = 0, x_2 = 1 \\ \mu_{10} & \text{如果} \quad x_1 = 1, x_2 = 0 \\ \mu_{11} & \text{如果} \quad x_1 = 1, x_2 = 1. \end{cases}$$

此条件期望函数可以写成 $(x_1, x_2, x_1x_2)'$ 的线性函数：

$$\mathbb{E}[y|x_1, x_2] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2,$$

其中，$\beta_0 = \mu_{00}$, $\beta_1 = \mu_{10} - \mu_{00}$, $\beta_2 = \mu_{01} - \mu_{00}$, $\beta_3 = \mu_{11} - \mu_{10} - \mu_{01} + \mu_{00}$。相应的饱和回归模型可以表示为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon.$$

在上式中，x_1 和 x_2 被称为主项，x_1x_2 被称为交互项。假设变量 y 表示工资，则 β_0 表示没有大学学历的男性的期望工资；β_1 衡量没有大学学历的女性与男性的期望工资的差异；β_2 衡量大学学历对男性期望工资的影响；$\beta_3 = (\mu_{11} - \mu_{10}) - (\mu_{01} - \mu_{00})$ 衡量的是大学学历对女性期望工资的影响与大学学历对男性期望工资的影响之间的差异。

最后，我们考虑二值虚拟变量和多值离散变量构成的饱和回归模型。假设 x_1 依然是表示性别的虚拟变量；x_2 表示受教育年限，可能的取值为 $\{0, 1, 2, \cdots, \tau\}$。相对于 x_1 和 x_2 的饱和模型可以表示为：

$$y = \beta_0 + \sum_{j=1}^{\tau} \beta_j d_{2j} + \gamma x_1 + \sum_{j=1}^{\tau} \delta_j (d_{2j} x_1) + \varepsilon,$$

其中，$d_{2j} = \mathbf{1}[x_2 = j]$，$j = 1, 2, \cdots, \tau$。通常，在回归分析中，若线性模型中含有交互项，

则相应的主项也应该出现在回归模型中。考虑如下去掉受教育年限主项的回归模型：

$$\mathrm{y} = \beta_0 + \gamma \mathrm{x}_1 + \sum_{j=1}^{\tau} \delta_j (\mathrm{d}_{2j}\mathrm{x}_1) + \varepsilon,$$

若 y 表示工资，这一模型只允许女性通过改变受教育年限影响工资收入，男性的工资收入和受教育年限无关，这与实际情况不符。

1.5 总结

本章详细介绍了条件期望、条件方差和线性投影的定义和主要性质，为后续金融计量学知识的学习奠定理论基础。通过本章的学习读者应该理解为什么条件期望在金融计量分析中的扮演着重要的角色，这涉及到 CEF 的分解性质和预测性质。此外，迭代期望律在金融计量方法理论性质的推导中经常被使用，本章给出了（简单）迭代期望律的详细证明，并在其他定理和性质的证明过程中反复使用此定律，希望读者能够掌握此定律并熟练使用。

本章的另外一个重要内容是线性投影。线性投影是条件期望函数的最优线性近似, 对变量 $\mathbf{x}$ 做投影得到的误差 e 与 $\mathbf{x}$ 正交，即，$\mathbb{E}[\mathbf{x}e] = \mathbf{0}$。我们在下一章将会看到最小二乘估计量满足的正则方程 (normal equation) 实际上是此正交条件的样本形式。因此，最小二乘估计量实际上估计的是线性投影系数，对线性投影系数的理解有助于读者深入理解线性回归。这里读者应该注意区分 CEF 的分解性质中出现的 ε 和线性投影误差 e，前者满足 $\mathbb{E}[\varepsilon|\mathbf{x}] = 0$；但是，通常情况下，$\mathbb{E}[e|\mathbf{x}] = 0$ 并不成立，除非 CEF 为线性形式。

本章的写作借鉴了如下优秀计量教材的相关章节：Angrist 和 Pischke 两位教授合著的 *Mostly harmless econometrics: An empiricist's companion* 的第三章；Hansen 教授的线上教材手稿 *Econometrics* 的第二章；Wooldridge 教授的 *Econometric Analysis of Cross Section and Panel Data* 的第二章。推荐想进一步深入学习本章内容的读者阅读。

这里借用 *Mostly harmless econometrics: An empiricist's companion* 第三章的标题"Making Regression Make Sense"来说明本章的目的：我们通过对总体性质的讨论，厘清回归分析感兴趣的参数，在利用具体的样本对这些参数进行估计之前，先明确在总体中这些参数的具体含义。

第一章习题

1. 证明：定理1.5的证明过程中关于 $\mathrm{Var}(\mathrm{y}|\mathbf{x})$ 的分解表达式的第三项为 0。
2. 试用 CEF 的分解性质证明定理1.6。
3. 证明：在假设 1.1 成立的条件下，定理1.7中定义的投影误差 e 满足 $\mathbb{E}[\mathbf{x}e] < \infty$。[2]
4. 证明：在定理1.7中，若 $\mathbf{x}$ 中含有常数项，则 $\mathbf{x}$ 与 e 不相关。
5. 证明：利用定理1.7中 (1) 的结论证明式 (1.5) 成立。

[2]不能利用 $\mathbb{E}[\mathbf{x}e] = \mathbf{0}$ 的结论。